AD 735858

ON ESTIMATING THE RELIABILITY OF A COMPONENT

SUBJECT TO SEVERAL DIFFERENT STRESSES

by

Satish Chandra

Technical Report No. 109
Department of Statistics ONR Contract

November 22, 1971

DEPARTMENT OF STATISTICS
Southern Methodist University

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
SOUTHERN METHODIST UNIVERSITY	UNCLASSIFIED
	2b. GROUP
	UNCLASSIFIED

3. REPORT TITLE

On estimating the reliability of a component subject to several different stresses.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Satish Chandra

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 22, 1971	55	18

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-68-A-0515	
b. PROJECT NO.	109
NR 042-260	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	OFFICE OF NAVAL RESEARCH

13. ABSTRACT

A great deal has been written concerning the estimation of the probability and testing of whether one of two random variables is stochastically larger than the other and its relationship to the estimation of reliability for stress-strength relationships. A more general problem is the estimation and testing of whether one of $N + 1$ random variables is simultaneously stochastically larger (smaller) than the others. An initial paper which deals with this problem for the special case $N = 2$ is that of D. R. Whitney (1951), "A Bivariate Extension of the U Statistic," where he provides a test function and discusses the asymptotic normality of the statistics proposed under the null hypothesis that all the random variables have the same distribution function.

In this dissertation, the problem of estimation of the probability of whether one of $N + 1$ mutually independent random variables, each having a continuous cumulative distribution function, is simultaneously stochastically larger (smaller) than the others has been considered. Parametric and nonparametric methods of estimation are discussed. Applications of the problem involve the estimation of reliability from stress-strength relationships, where a component is subject to several stresses (several strengths) whereas its strength (stress) is a single random variable.

Minimum variance unbiased estimates are provided in certain cases, whereas in some cases the maximum likelihood estimates are given. For a more general situation, where only continuity of the cumulative distribution functions of $N + 1$ random variables is assumed, a statistic W, somewhat similar to the Mann-Whitney U statistic, based on ranks of the pooled sample values from all the random variables, has been suggested and shown to have certain optimum properties; i.e., it's an unbiased & consistent est.

DD FORM 1473 (PAGE 1)
I NOV 65

S/N 0101-807-6801

CHAPTER I

INTRODUCTION

In the past few years, many authors (e.g., Z. W. Birnbaum, McCarty, Church, Harris, VanDantzig, Govindarajulu, Mann, Whitney, Mazumdar, Owen, Sen, etc.) have attempted to estimate and give the confidence bounds for the reliability of a component using the probability arguments of a certain physical model of failure. According to this model a component fails if at any moment the applied stress (or load), say Y, is greater than the resistance, say X. That is, the problem here is to find an estimate of the probability that X is less than Y, where X and Y are both random variables, having some known or unknown probability structure, and to find the confidence limits on this estimate of the probability. An extensive amount of work has been done on this problem by the above mentioned authors and many others, and some, for example, Mann and Whitney [9], Mazumdar [10], Church and Harris [3] etc. have given practical uses for the results.

The question now arises of what to do if at any moment the applied stress (or load or force) may not be measured in terms of a single random variable, but can be measured in terms of several random variables, say $X_1, X_2, \ldots, X_N$, and the resistance is still a single random variable, say Y. These random variables $X_1, X_2, \ldots, X_N, Y$ may have a known (specified) probability structure to a certain extent or the structure may be completely

unspecified except that in this latter case independence of the X's is assumed. In some cases, it might be possible to arrive at a suitable mathematical model which allows one to compute deterministically the "over-all stress" at different points of time corresponding to a given set of "initial stresses," and initial conditions, but it may not be possible in all the cases, especially when all the initial stresses are random variables. For example, the over-all stress, say X, may be a suitable linear combination of all the initial stresses $X_1, \ldots, X_N$, with $X_1, \ldots, X_N$ being random variables. Thus, writing $X = \sum_{i=1}^{N} C_i X_i$, C_i's being real constants, the problem is to estimate C_i's, the parameters involved in order to estimate the over-all stress X. A robust way to estimate these constants is still under consideration. Some authors have attempted to estimate these constants using the mixture of probability distributions techniques in some simple cases, for example when $N = 2$ and X_1, X_2 are independent normal random variables. The problem is even more complicated if the over-all stress is a nonlinear combination of the initial stresses.

In many cases, a component does not fail if all the different stresses $X_1, \ldots, X_N$ are simultaneously less than the resistance Y. Then, a measure of reliability of a component subject to several different stresses during a given period [0,T] is taken to be the probability that all the different stresses are simultaneously less than the resistance during the entire interval, i.e.

(1.1) $p = \text{Prob} \{X_1 < Y, X_2 < Y, \ldots, X_N < Y\}$.

Church and Harris [3] and Mazumdar [10] considered the example of

missile flights, where the several different stresses are the propulsive force, angles of elevation, changes in atmospheric condition, and so on. They assumed that all these different stresses have known distributions, and a mathematical model is available to compute the over-all stress, and also the distribution of over-all stress is known. However, these assumptions might not be true in a given physical situation.

Another similar problem of interest is that where during a given time period [0,T] the stress or force could be measured in terms of one random variable, say X whereas a component might consist of several different objects subject to different resistances. For example, suppose the electric current is supplied to a component consisting of several different transistors each of which has a different capacity to resist the current. Thus, the component breaks if the current supplied to the component exceeds the capacity of any of the transistors. Hence, denoting the stress or force of the component by X, and the several different resistances by $Y_1, Y_2, \ldots, Y_N$, the component does not fail if the applied force or stress X is simultaneously less than the several different resistances. A measure of reliability of a component subject to single force and various resistances during a given period [0,T] can be taken to be the Probability that the applied force is simultaneously less than the several different resistances during the entire interval, i.e.

(1.2) $\qquad q = \text{Prob } \{X < Y_1, \ldots, X < Y_N\}.$

where $X, Y_1, \ldots, Y_N$ are all random variables.

The principal objective of the material presented here is to estimate p and q defined above.

Chapter II consists of parametric estimation of p, assuming that $X_1,\ldots,X_N,Y$ have certain probability structures.

Chapter III consists of a non-parametric approach to estimate p for a much more general situation, i.e., $X_1,\ldots,X_N,Y$ are all independent random variables having different absolutely continuous distribution functions. A statistic is proposed and shown to estimate p unbiasedly. The variance of this statistic is computed and used to show that this estimate of p is consistent.

Chapter IV is devoted to the estimation of q. Analogous results to Chapter II and III are given and some more results of practical importance are included.

Finally, Chapter V consists of a summary and suggestions for further research.

CHAPTER II

PARAMETRIC ESTIMATION

2.1 Introduction

Owen, Craswell and Hanson [12], Govindarajula [5], Church and Harris
[3], etc. have given maximum likelihood estimates and confidence intervals
for Prob $\{X < Y\}$ under the assumptions that X and Y have a bivariate nor-
mal distribution with some parameters known, and for the case when X and
Y are independent, and when they are paired. Mazumdar [10] has given the
minimum variance unbiased estimate for Prob $\{X < Y\}$ when X is assumed to
be normally distributed with known mean and known variance; also Y is
normally distributed with unknown mean, known or unknown variance and X,Y
being independent. He also derived interval estimates for these cases.

In this chapter, we shall consider the estimation of the probability
that N random variables $X_1, \ldots, X_N$ are simultaneously less than a random
variable Y, where these random variables are mutually independent.

2.2 Useful Lemmas

Let $X_1, \ldots, X_N, Y$ be $(N + 1)$ mutually independent random variables with
continuous cumulative distribution functions (c.d.f.'s) $F_1, F_2, \ldots, F_N, G$
respectively and let

(2.1) $T = \max(X_1, \ldots, X_N)$

and have a c.d.f. H. Also, let

-5-

$$p = \text{Prob}\{X_1 < Y, \ldots, X_N < Y\}.$$

Then, the problem of estimating p is simplified using the following lemmas.

Lemma 2.1

$$P(X_1 < Y, \ldots, X_N < Y) = P(T < Y).$$

Lemma 2.2

$$H(t) = P(T \leq t) = \prod_{i=1}^{N} F_i(t).$$

Proof

$$\begin{aligned}
H(t) &= P(T \leq t) = P(\max(X_1, \ldots, X_N) \leq t) \\
&= P(X_1 \leq t, \ldots, X_N \leq t) \\
&= P(X_1 \leq t) \ldots \ldots P(X_N \leq t) \\
&= \prod_{i=1}^{N} F_i(t).
\end{aligned}$$

Since $H(t)$ is a function of continuous c.d.f.'s $H(t)$ itself is continuous and so the probability of getting ties from the c.d.f. H is zero. Moreover, the random variable T is independent of the random variable Y.

Now, suppose $F_1 = F_2 = F_N = F$, i.e., the random variables $X_1, \ldots, X_N$ have the same continuous c.d.f. F. Also, F has the density f and Y has the continuous c.d.f. G with the density g, and all these random variables are mutually independent. Then, the density function of the random variable $T = \max(X_1, \ldots, X_N)$ is given as

$$(2.2) \qquad\qquad h(t) = N(F(t))^{N-1} f(t)$$

and thus,

Lemma 2.3

$$(2.3) \qquad p = \int_{-\infty}^{\infty} [F(y)]^N \, dG(y) \quad.$$

Proof

$$p = P(T < Y)$$

$$= E_Y P(T < Y|Y)$$

$$= \int_{-\infty}^{\infty} [F(y)]^N \, dG(y) \ .$$

Lemma 2.4

If $F_1 = F_2 = F_N = G$, then

$$p = \frac{1}{N+1} \ .$$

Proof

In this case, from Equation (2.3)

$$p = \int_{-\infty}^{\infty} F^N(y) \, dF(y)$$

$$= \int_0^1 u^N \, du = \frac{1}{N+1} \ .$$

2.3 Estimation of p when $X_1, X_2, \ldots, X_N, Y$ are all independent normal

Let Φ and ϕ denote the cumulative distribution frunction and the density of a standard normal random variable respectively, i.e.,

(2.4) $\phi(y) = \dfrac{1}{(2\pi)^{1/2}} \ e^{-1/2 \ y^2}$ $-\infty < y < \infty$

and

(2.5) $\Phi(x) = \displaystyle\int_{-\infty}^{x} \phi(y) \, dy.$

Further, let the different stresses $X_1, \ldots, X_N$ be all independent and

identically distributed normal random variable with mean μ, variance σ^2. Also, let Y be normal with mean ν, variance τ^2 and independent of all the X's. Then by lemma 2.3,

$$p = \int_{-\infty}^{\infty} [\int_{-\infty}^{y} \frac{1}{(2\pi)^{1/2}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx]^N \frac{1}{(2\pi)^{1/2}\tau} e^{-\frac{1}{2\tau^2}(y-\nu)^2} dy$$

$$= \int_{-\infty}^{\infty} \phi^N(\frac{\nu + \tau z - \mu}{\sigma}) \phi(z) dz$$

$$= \int_{-\infty}^{\infty} \phi^N \left(\frac{(\frac{\tau^2}{\sigma^2 + \tau^2})^{1/2} z + (\frac{1}{\sigma^2 + \tau^2})^{1/2}(\nu - \mu)}{(1 - \frac{\tau^2}{\sigma^2 + \tau^2})^{1/2}} \right) \phi(z) dz$$

$$= \int_{-\infty}^{\infty} \phi^N (\frac{\rho^{1/2} z + \tilde{H}}{(1-\rho)^{1/2}}) \phi(z) dz$$

$$= F_N(\tilde{H} ; \rho) \quad , \text{say}$$

where $\rho = \frac{\tau^2}{\sigma^2 + \tau^2}$

$$\tilde{H} = -\frac{(\nu - \mu)}{(\sigma^2 + \tau^2)^{1/2}} .$$

The function $F_N(\tilde{H},\rho)$ has been tabulated by Shanti S. Gupta [6] for given values of $\rho(\geq 0)$, N and $\tilde{H}$. $F_N(\tilde{H},\rho)$ is the probability that each of N standarized normal variables with equal correlation ρ will not exceed $\tilde{H}$. This is the case in the present problem, since

$$P\{X_1 < Y,\ldots,X_N < Y\} = P\{X_1 - Y < 0,\ldots,X_N - Y < 0\}$$

$$= P\left\{ \frac{(X_1-Y) - (\mu-\nu)}{(\sigma^2 + \tau^2)^{1/2}} < \frac{\nu - \mu}{(\sigma^2 + \tau^2)^{1/2}} ,\ldots, \frac{(X_N-Y) - (\mu-\nu)}{(\sigma^2 + \tau^2)^{1/2}} \right.$$

$$\left. < \frac{\nu - \mu}{(\sigma^2 + \tau^2)^{1/2}} \right\} .$$

For all $i = 1,\ldots,N$, the random variable

$$\frac{(X_i - Y) - (\mu - \nu)}{(\sigma^2 + \tau^2)^{1/2}}$$

is a standarized normal random variable. Moreover, for $i \neq j$, the correlation coefficient between

$$\frac{(X_i - Y) - (\mu - \nu)}{(\sigma^2 + \tau^2)^{1/2}} \quad \text{and} \quad \frac{(X_j - Y) - (\mu - \nu)}{(\sigma^2 + \tau^2)^{1/2}} \quad \text{is} \quad \frac{\tau^2}{\sigma^2 + \tau^2} .$$

$F_N(\tilde{H}, \rho)$ also gives the probability that the minimum of a set of N equally correlated standardized normal random variables exceeds $-\tilde{H}$.

Thus, if μ, ν, σ^2 and τ^2 are known we can evaluate p using Gupta's table. However if μ, ν, σ^2, τ^2 are unknown, or some of them are unknown, we can use their maximum likelihood estimates. If μ, ν, σ^2 and τ^2 are unknown, and if the samples

$$x_{i1}, \ldots, x_{in_i} \qquad i = 1,\ldots,N,$$
$$y_1, \ldots, y_m$$

are available, let

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} , \qquad i = 1,\ldots,N ,$$

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 ,$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} \bar{x}_i ,$$

$$s^2 = \frac{\sum_{i=1}^{N} n_i s_i^2}{\sum_{i=1}^{N} n_i} ,$$

$$\bar{y} = \frac{1}{m} \sum_{k=1}^{m} y_k \quad ,$$

$$(s')^2 = \frac{1}{m} \sum_{k=1}^{m} (y_k - \bar{y})^2 \quad .$$

Then, the maximum likelihood estimate of p, in view of Zehna's [18] result that the maximum likelihood estimate remains invariant although p is not a one-to-one function of μ, ν, ζ^2 and τ^2, is

$$\hat{p} = F_N(\hat{\tilde{H}}, \hat{\rho}) \quad ,$$

where

$$\hat{\rho} = \frac{s^2}{s^2 + (s')^2} \quad ,$$

$$\hat{\tilde{H}} = \frac{\bar{y} - \bar{x}}{(s^2 + (s')^2)^{1/2}} \quad .$$

This estimate, again, can be evaluated using Gupta's table [6], however the properties of this estimate are extremely difficult to explore.

2.4 Minimum Variance Unbiased Estimation of p in Some Cases

Let X, Y, Z be independent normal with means μ_x, μ_y and μ_z respectively and equal variance σ^2. Then we want to estimate

$$p = P\{X < Z, \quad Y < Z\}$$
$$= P\{X - Z < 0, \quad Y - Z < 0\}$$

Now, (X - Z, Y - Z) is bivariate normally distributed with means $(\mu_x - \mu_y,$ $\mu_y - \mu_z)$ and covariance matrix

$$\sigma^2 \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Thus,

$$p = P\left\{ \frac{(X - Z) - (\mu_x - \mu_z)}{(2)^{1/2}\,\sigma} < \frac{\mu_z - \mu_x}{(2)^{1/2}\,\sigma} \;, \right.$$

$$\left. \frac{(Y - Z) - (\mu_y - \mu_z)}{(2)^{1/2}\,\sigma} < \frac{\mu_z - \mu_y}{(2)^{1/2}\,\sigma} \right\} \;.$$

Now,

$$\left(\frac{(X - Z) - (\mu_x - \mu_z)}{(2)^{1/2}\,\sigma} \;, \; \frac{(Y - Z) - (\mu_y - \mu_z)}{(2)^{1/2}\,\sigma} \right)$$

is bivariate normal with zero means, unit variances and correlation coefficient equal to 1/2. Thus

$$p = B\left(\frac{\mu_z - \mu_x}{(2)^{1/2}\sigma} \;, \; \frac{\mu_z - \mu_y}{(2)^{1/2}\sigma} \;, \; \frac{1}{2} \right) \;,$$

where

$$(2.6) \quad B(h,K,\rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \int_{-\infty}^{h}\int_{-\infty}^{K} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} \; dx\,dy \;.$$

D. B. Owen [11] considered evaluation of the function $B(h,K,\rho)$ using the related function $T(h,a)$ defined by

$$T(h,a) = \frac{1}{2\pi} \int_{0}^{a} \frac{\exp\{-\frac{1}{2}h^2(1 + x^2)\}}{1 + x^2} \; dx \;.$$

He derived various relations connecting $B(h, K, \rho)$ and $T(h,a)$. The function $T(h,a)$ and its differences are tabulated by Owen in [11]. Using his tables and relations one can obtain the volume under a bivariate surface over any polygon.

Assume that samples $(x_1,\ldots,x_\ell)$, $(y_1,\ldots,y_m)$ and $(z_1,\ldots,z_n)$ are available. If μ_x, μ_y, μ_z, and σ are unknown, we substitute their maximum

likelihood estimates to obtain the maximum likelihood estimate of p as

$$\tilde{p} = B\left(\frac{\bar{z} - \bar{x}}{(2)^{1/2}s}, \frac{\bar{z} - \bar{y}}{(2)^{1/2}s}, \frac{1}{2}\right),$$

where

$$\bar{x} = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i,$$

$$\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_j,$$

$$\bar{z} = \frac{1}{n} \sum_{k=1}^{n} z_k,$$

and

$$s^2 = \frac{1}{\ell+m+n}\left[\sum_1^{\ell}(x_i - \bar{x})^2 + \sum_1^m(y_j - \bar{y})^2 + \sum_1^n(z_k - \bar{z})^2\right].$$

Theorem 2.1

If the means μ_x, μ_y and variance σ^2 are known, but μ_z is unknown and the data $(z_1,\ldots,z_n)$ on Z are available, then the minimum variance unbiased estimate of p is given as

$$\hat{p} = B\left(\frac{\bar{z} - \mu_x}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{\bar{z} - \mu_y}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{n-1}{2n-1}\right),$$

where

$$\bar{z} = \frac{1}{n} \sum_{k=1}^{n} z_k.$$

Moreover, this estimate is asymptotically equivalent to the maximum likelihood estimate.

Proof

Let W and V be independent normal random variables, independent of Z, with means 0 and variance σ^2, and let

$$I = \begin{cases} 1 & \text{if } -z_1 + W < -\mu_x \ , \quad -z_1 + V < -\mu_y \\ 0 & \text{otherwise} \end{cases} \ ,$$

then

$$E(I) = P(-z_1 + W < -\mu_x \ , \quad -z_1 + V < -\mu_y)$$

$$= P(\frac{-z_1 + W + \mu_z}{(2)^{1/2}\sigma} < \frac{\mu_z - \mu_x}{(2)^{1/2}\sigma} \ , \quad \frac{-z_1 + V + \mu_z}{(2)^{1/2}\sigma} < \frac{\mu_z - \mu_y}{(2)^{1/2}\sigma})$$

$$= B(\frac{\mu_z - \mu_x}{(2)^{1/2}\sigma} \ , \quad \frac{\mu_z - \mu_y}{(2)^{1/2}\sigma} \ , \quad \frac{1}{2})$$

$$= p \ .$$

Since $(\dfrac{-z_1 + W + \mu_z}{(2)^{1/2}\sigma} \ , \quad \dfrac{-z_1 + V + \mu_z}{(2)^{1/2}\sigma})$ is bivariate normal with zero

means, unit variances and correlation coefficient equal to $1/2$.

Since $\bar{z}$ is a complete sufficient statistic, it follows from the Rao-Blackwell and Lehmann-Scheffe$'$ theorems that the conditional expectation $E(I|\bar{z})$ is the minimum variance unbiased (m.v.u.) estimate of p.

Let $N_p(\mu, \Sigma)$ denote the p-variate normal distribution with mean vector μ and variance-covariance matrix Σ; then,

$$\begin{pmatrix} \bar{z} \\ -z_1 + W \\ -z_1 + V \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu_z \\ -\mu_z \\ -\mu_z \end{pmatrix} , \begin{pmatrix} \frac{\sigma^2}{n} & -\frac{\sigma^2}{n} & -\frac{\sigma^2}{n} \\ -\frac{\sigma^2}{n} & 2\sigma^2 & \sigma^2 \\ -\frac{\sigma^2}{n} & \sigma^2 & 2\sigma^2 \end{pmatrix} \right) ,$$

where $\sim$ denotes "distributed" as.

Now, using 8a.2(v), Page 441, Rao [14], the conditional distribution of $(-z_1 + W \ , \quad -z_1 + V)$ given $\bar{z}$ is

$$N_2 \left(\begin{pmatrix} -\bar{z} \\ -\bar{z} \end{pmatrix} , \begin{pmatrix} 2\sigma^2 - \frac{\sigma^2}{n} & \sigma^2 - \frac{\sigma^2}{n} \\ \sigma^2 - \frac{\sigma^2}{n} & 2\sigma^2 - \frac{\sigma^2}{n} \end{pmatrix} \right) ,$$

and the conditional correlation coefficient is $\frac{n-1}{2n-1}$.

Thus,

$$\hat{p} = E(I|\bar{z}) = P\{-z_1 + W < -\mu_x \, , \quad -z_1 + V < -\mu_y|\bar{z}\}$$

$$P\left\{\frac{-z_1 + W + \bar{z}}{\sigma(2-\frac{1}{n})^{1/2}} < \frac{\bar{z} - \mu_x}{\sigma(2-\frac{1}{n})^{1/2}} \, , \quad \frac{-z_1 + V + \bar{z}}{\sigma(2-\frac{1}{n})^{1/2}} < \frac{\bar{z} - \mu_y}{\sigma(2-\frac{1}{n})^{1/2}} \, \Big| \bar{z}\right\}$$

i.e.

$$\hat{p} = B\left(\frac{\bar{z} - \mu_x}{\sigma(2-\frac{1}{n})^{1/2}} \, , \quad \frac{\bar{z} - \mu_y}{\sigma(2-\frac{1}{n})^{1/2}} \, , \quad \frac{n-1}{2n-1}\right)$$

and the maximum likelihood estimate of p in this situation is

$$\tilde{p} = B\left(\frac{\bar{z} - \mu_x}{(2)^{1/2}\sigma} \, , \quad \frac{\bar{z} - \mu_y}{(2)^{1/2}\sigma} \, , \quad \frac{1}{2}\right) \, .$$

Thus, as n becomes large, the minimum variance unbiased estimate approaches in the limit to the maximum likelihood estimate.

Theorem 2.2

If the mean μ_z and variance σ^2 are known but the means μ_x and μ_y are unknown, and the data $(x_1,\ldots,x_\ell)$, $(y_1,\ldots,y_m)$ on X and Y respectively are available, then the m.v.u. estimate of p is given by

$$\hat{p} = B\left(\frac{\mu_z - \bar{x}}{\sigma(2-\frac{1}{\ell})^{1/2}} \, , \quad \frac{\mu_z - \bar{y}}{\sigma(2-\frac{1}{m})^{1/2}} \, , \quad \frac{1}{[(2-\frac{1}{\ell})(2-\frac{1}{m})]^{1/2}}\right)$$

where

$$\bar{x} = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i \, ,$$

$$\bar{y} = \frac{1}{m} \sum_{j=1}^{m} y_j \, .$$

Moreover, as ℓ and m become large, this estimate approaches in the limit to the maximum likelihood estimate.

Proof

Using the same technique as in the proof of Theorem 2.1, let W be a normal random variable, independent of X and Y, with mean 0 and variance σ^2, and let

$$I = \begin{cases} 1 & \text{if } x_1 + W < \mu_z \ , \ y_1 + W < \mu_z \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E(I) = P(x_1 + W < \mu_z \ , \ y_1 + W < \mu_z)$$

$$= P\left(\frac{x_1 + W - \mu_x}{(2)^{1/2}\sigma} < \frac{\mu_z - \mu_x}{(2)^{1/2}\sigma} \ , \ \frac{y_1 + W - \mu_y}{(2)^{1/2}\sigma} < \frac{\mu_z - \mu_y}{(2)^{1/2}\sigma}\right)$$

$$= B\left(\frac{\mu_z - \mu_x}{(2)^{1/2}\sigma} \ , \ \frac{\mu_z - \mu_y}{(2)^{1/2}\sigma} \ , \ \frac{1}{2}\right)$$

$$= p \ .$$

Since $(\bar{x}, \bar{y})$ is a complete sufficient statistic, it follows from the Rao-Blackwell and Lehmann-Scheffé theorem that the conditional expectation $E(I|\bar{x}, \bar{y})$ is the m.v.u. estimate of p. Now

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ x_1 + W \\ y_1 + W \end{pmatrix} \sim N_4 \left(\begin{pmatrix} \mu_x \\ \mu_y \\ \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{\ell} & 0 & \frac{\sigma^2}{\ell} & 0 \\ 0 & \frac{\sigma^2}{m} & 0 & \frac{\sigma^2}{m} \\ \frac{\sigma^2}{\ell} & 0 & 2\sigma^2 & \sigma^2 \\ 0 & \frac{\sigma^2}{m} & \sigma^2 & 2\sigma^2 \end{pmatrix} \right).$$

Using 8a.2(V) Page 441, Rao [14], the conditional distribution of $(x_1 + W, y_1 + W)$ given $(\bar{x}, \bar{y})$ is

$$N_2 \left(\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} 2\sigma^2 - \frac{\sigma^2}{\ell} & \sigma^2 \\ \sigma^2 & 2\sigma^2 - \frac{\sigma^2}{m} \end{pmatrix} \right),$$

and conditional correlation coefficient is $\dfrac{1}{[(2 - \frac{1}{\ell})(2 - \frac{1}{m})]^{1/2}}$

and hence the m.v.u. estimate of p is

$$\hat{p} = P\{x_1 + W < \mu_z \quad , \quad y_1 + W < \mu_z | \bar{x}, \bar{y}\}$$

$$= P\{\frac{x_1 + W - \bar{x}}{\sigma(2 - \frac{1}{\ell})^{1/2}} < \frac{\mu_z - \bar{x}}{\sigma(2 - \frac{1}{\ell})^{1/2}} \quad , \quad \frac{y_1 + W - \bar{y}}{\sigma(2 - \frac{1}{m})^{1/2}} < \frac{\mu_z - \bar{y}}{\sigma(2 - \frac{1}{m})^{1/2}} \}$$

$$= B(\frac{\mu_z - \bar{x}}{\sigma(2 - \frac{1}{\ell})^{1/2}} \quad , \quad \frac{\mu_z - \bar{y}}{\sigma(2 - \frac{1}{m})^{1/2}} \quad , \quad \frac{1}{[(2 - \frac{1}{\ell})(2 - \frac{1}{m})]^{1/2}})$$

and this estimate is asymptotically equivalent to the maximum likelihood

estimate as ℓ, m become large.

CHAPTER III

NON-PARAMETRIC ESTIMATION

3.1 Introduction

Mann and Whitney [9] proposed a U statistic which estimates unbiasedly the probability that X is less than Y where X and Y are both random variables. The U statistic is based on the ranks of observations in the pooled samples of X's and Y's. It is also used for testing the hypothesis that X's and Y's come from the same population versus the alternative that X is stochastically smaller than Y. Many properties of this estimate have been discussed by numerous authors.

Here, a statistic is proposed to estimate the probability that all of the N random variables X_1, X_2,...,X_N are simultaneously less than a random variable Y. It is assumed that X's and Y's have continuous cumulative distribution functions (c.d.f.'s) and that all are mutually independent. Some properties of the proposed estimators are also discussed.

We shall first start with the simple case when N = 2. Let X, Y, Z be mutually independent random variables with continuous c.d.f.'s F_1, F_2 and G respectively, and let

(3.1) $p = \text{Prob}\{X < Z$, $Y < Z\}$.

Again let

(3.2) $x_1,...,x_\ell$; $y_1,...,y_m$; $z_1,...,z_n$

be samples of X, Y and Z respectively. Furthermore, we arrange the samples

-17-

of ℓ x's, m y's and n z's in ascending order and let W count the number of times an x and a y precedes a z, i.e.,

W = number of triplets (x_i, y_j, z_k) such that $x_i < z_k$ and $y_j < z_k$

$$i = 1,\ldots,\ell$$

$$j = 1,\ldots,m$$

$$k = 1,\ldots,n$$

then $\hat{p} = \dfrac{W}{\ell mn}$ can be used to estimate the parameter p. The properties of this particular estimator will be explored in the following sections.

3.2 Estimation of p and an Example

Let samples (3.2) be available and let

$$a_{ik} = \begin{cases} 1 & \text{if } x_i < z_k \\ 0 & \text{otherwise}, \end{cases}$$

(3.3)

$$b_{jk} = \begin{cases} 1 & \text{if } y_j < z_k \\ 0 & \text{otherwise}. \end{cases}$$

Furthermore, let

$$U_{ijk} = a_{ik}b_{jk},$$

then

$$W = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} a_{ik}b_{jk} \ .$$

Lemma 3.1

$\hat{p} = \dfrac{W}{\ell mn}$ is an unbiased estimate of p.

Proof

$$E(\hat{p}) = \frac{1}{\ell mn} \sum_i \sum_j \sum_k E(U_{ijk})$$

$$= \frac{1}{\ell mn} \sum_i^{\ell} \sum_j^{m} \sum_k^{n} p = p.$$

Since $U_{ijk} = 1$ if and only if (iff) $x_i < z_k$ and $y_j < z_k$ and so

$$E(U_{ijk}) = P(x_i < z_k , y_j < z_k) = p .$$

Example 3.1

Let $\ell = 4$, $m = 3$, $n = 2$ and consider the observations

$$x_2 < y_3 < z_2 < x_1 < x_4 < y_1 < z_1 < x_3 < y_2 .$$

There are 6 values of (x,y) which are less than z_1.

There is 1 value of (x,y) which is less than z_2.

Thus the experimental value of W is

$$w = 6 + 1 = 7 ,$$

and

$$\hat{p} = \frac{7}{24} .$$

Also

$$a_{11} = 1, \ a_{21} = 1, \ a_{31} = 0, \ a_{41} = 0, \ a_{12} = 0, \ a_{22} = 1, \ a_{32} = 0,$$
$$a_{42} = 0,$$

and

$$b_{11} = 1, \ b_{21} = 0, \ b_{31} = 1, \ b_{12} = 0, \ b_{22} = 0, \ b_{32} = 1,$$

so that

$$\hat{p} = \frac{1}{\ell mn} \sum_{i=1}^{4} \sum_{j=1}^{3} \sum_{k=1}^{2} a_{ik} b_{jk} = \frac{7}{24} .$$

3.3 Properties of $\hat{p}$

In this section we shall compute the variance of $\hat{p}$ and use this variance to show that $\hat{p}$ is a consistent estimate of p. Let

$$T = \max(X,Y),$$

and H denote the c.d.f. of T which is continuous because H is a function of c.d.f.'s of X and Y and the c.d.f.'s of X and Y are continuous by as-

sumption. Then

$$p = \text{Prob}\{X < Z, \quad Y < Z\}$$

$$= \text{Prob}\{T < Z\}$$

$$= E_z \text{Prob}\{T < Z|Z\}$$

(where E denotes the expectation operator)

$$= \int_{-\infty}^{\infty} dG(z) \int_{-\infty}^{z} dH(t)$$

$$= \int_{-\infty}^{\infty} H(z) dG(z) ,$$

i.e.,

$$(3.4) \qquad p = \int_{-\infty}^{\infty} H(z) dG(z) .$$

Also,

$$1-p = P\{Z < T\}$$

$$= E_T\{Z < T|T\}$$

$$= \int_{-\infty}^{\infty} G(t) dH(t) .$$

Thus,

$$(3.5) \quad 1-p = \int_{-\infty}^{\infty} G(t) dH(t) .$$

Now, let

$$\psi^2 = \sigma^2_{H(z)} = E_z [H(z) - E_z(H(z))]^2$$

$$= \int_{-\infty}^{\infty} [H(z) - \int_{-\infty}^{\infty} H(z) dG(z)]^2 dG(z)$$

$$= \int_{-\infty}^{\infty} (H(z) - p)^2 \, dG(z) \quad \text{using (3.4)}.$$

Thus, it follows that

$$(3.6) \qquad \int_{-\infty}^{\infty} H^2(z) \, dG(z) = \psi^2 + p^2.$$

Again, let

$$\chi^2 = \sigma^2_{G(T)} = E_T[G(T) - E_T(G(T))]^2$$

$$= \int_{-\infty}^{\infty} [G(T) - \int_{-\infty}^{\infty} G(T) \, dH(T)]^2 \, dH(T)$$

$$= \int_{-\infty}^{\infty} [G(T) - 1 + p]^2 \, dH(T) \ , \ \text{using (3.5)}$$

Hence,

$$(3.7) \qquad \int_{-\infty}^{\infty} (1 - G(t))^2 \, dH(t) = \chi^2 + p^2.$$

We shall use the above relations in deriving the variance of $\hat{p}$.

Also used are

$$W = \sum_i^\ell \sum_j^m \sum_k^n U_{ijk} \ ,$$

and

$$E(W) = \ell m n p \ .$$

Theorem 3.1

The variance of W is given by

$$\text{Var}(W) = \ell m n [(\ell m - 1)\psi^2 + (n-1)(\ell + m - 1)\chi^2 + (\ell + m - 1)p$$

$$+ \{(n-1)(\ell + m - 2) - 1\}p^2] \ .$$

Proof

$$Var(W) = E(W^2) - (E(W))^2$$

$$= E(W^2) - (\ell mnp)^2 ,$$

so that we need to compute $E(W^2)$.

$$E(W^2) = E(\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} U_{ijk})^2$$

$$= E(\sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{k=1}^{n} \sum_{k'=1}^{n} U_{ijk} U_{i'j'k'}) .$$

To compute this expected value we shall consider the following cases.

1) $i = i'$, $j = j'$, $k = k'$.

Here $U_{ijk} U_{i'j'k'} = U_{ijk}^2 = 1$ if and only if (iff)

$x_i < z_k$ and $y_j < z_k$.

Thus,

$$E(U_{ijk} U_{i'j'k'}) = P(x_i < z_k , y_j < z_k) = p.$$

ii) $i = i'$, $j = j'$, $k \neq k'$.

Here $U_{ijk} U_{ijk'} = 1$ iff $\max(x_i,y_j) < z_k$ and $\max(x_i,y_j) < z_{k'}$

i.e. iff $t < z_k$ and $t < z_{k'}$,

where $t = \max(x_i,y_j)$,

thus

$$E(U_{ijk} U_{ijk'}) = P(t < z_k \text{ and } t < z_{k'})$$

$$= E_t P(z_k > t , z_{k'} > t | t)$$

$$= \int_{-\infty}^{\infty} dH(t) \int_{t}^{\infty} dG(z_k) \int_{t}^{\infty} dG(z_{k'})$$

since for $k \neq k'$, z_k, $z_{k'}$ are independent

$$= \int_{-\infty}^{\infty} (1 - G(t))^2 dH(t)$$

$$= \chi^2 + p^2 \quad \text{using (3.7)}.$$

iii) $i = i'$, $j \neq j'$, $k \neq k'$.

Here $U_{ijk} U_{ij'k'} = 1$ iff $\max(x_i, y_j) < z_k$ and $\max(x_i, y_{j'}) < z_{k'}$

i.e. iff $t_1 < z_k$ and $t_2 < z_{k'}$,

where $t_1 = \max(x_i, y_j)$

$t_2 = \max(x_i, y_{j'})$.

Note that t_1 may be equal to t_2 with nonzero probability, i.e. when t_1 and t_2 are the same observations. Thus

$$E(U_{ijk} U_{ij'k'}) = P(t_1 < z_k, \ t_2 < z_{k'} | t_1 = t_2 = t)$$

$$+ P(t_1 < z_k, \ t_2 < z_{k'} | t_1 \neq t_2)$$

$$= \chi^2 + p^2 + p^2$$

$$= \chi^2 + 2p^2 ,$$

using case (ii) and noting that when $t_1 \neq t_2$ and $k \neq k'$,

$$P(t_1 < z_k, \ t_2 < z_{k'}) = P(t_1 < z_k) P(t_2 < z_{k'}) = p^2.$$

iv) $i = i'$, $j \neq j'$, $k = k'$.

Here $U_{ijk} U_{ij'k} = 1$ iff $\max(x_i, y_j) < z_k$ and $\max(x_i, y_{j'}) < z_k$

i.e. iff $t_1 < z_k$ and $t_2 < z_k$,

where $t_1 = \max(x_i, y_j)$ and $t_2 = \max(x_i, y_{j'})$.

Again, t_1 may be equal to t_2 with nonzero probability. Thus

$$E(U_{ijk} \, U_{ij'k}) = P(t_1 < z_k \, , \, t_2 < z_k | t_1 = t_2 = t)$$

$$+ \, P(t_1 < z_k \, , \, t_2 < z_k | t_1 \neq t_2)$$

$$= P(t < z_k) + E_{z_k} P(t_1 < z_k \, , \, t_2 < z_k | z_k)$$

$$= p + \int_{-\infty}^{\infty} dG(z_k) \int_{-\infty}^{z_k} dH(t_1) \int_{-\infty}^{z_k} dH(t_2)$$

$$= p + \int_{-\infty}^{\infty} (H(z_k))^2 dG(z_k)$$

$$= p + \psi^2 + p^2 \qquad \text{using (3.6)}.$$

v) $i \neq i'$, $j = j'$, $k = k'$.

By symmetry as in case (iv) we have

$$E(U_{ijk} \, U_{i'jk}) = p + \psi^2 + p^2 \, .$$

vi) $i \neq i'$, $j = j'$, $k \neq k'$.

By symmetry as in case (iii), we have

$$E(U_{ijk} \, U_{i'jk'}) = \chi^2 + 2p^2 \, .$$

vii) $i \neq i'$, $j \neq j'$, $k = k'$

Here $U_{ijk} \, U_{i'j'k} = 1$ iff $\max(x_i, y_j) < z_k$ and $\max(x_{i'}, y_{j'}) < z_k$

i.e. iff $t_1 < z_k$ and $t_2 < z_k$,

where $t_1 = \max(x_i, y_j)$ and $t_2 = \max(x_{i'}, y_{j'})$.

In this case $t_1 \neq t_2$ with probability 1. Thus

$$E(U_{ijk} \, U_{i'j'k}) = P(t_1 < z_k \, , \, t_2 < z_k)$$

$$= \psi^2 + p^2 \qquad \text{using the result from case (iv)}.$$

viii) $i \neq i'$, $j \neq j'$, $k \neq k'$.

   In this case U_{ijk} , $U_{i'j'k'}$ are independent, so that

   $E(U_{ijk} U_{i'j'k'})$ = p^2.

Hence,

$$E(W^2) = \sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{k=1}^{n} \sum_{k'=1}^{n} E(U_{ijk} U_{i'j'k'})$$

$$= \ell m n p + \ell m n (n-1)(\chi^2 + p^2) + \ell m (m-1) n (n-1)(\chi^2 + 2p^2)$$

$$+ \ell m (m-1) n (p + \psi^2 + p^2) + \ell(\ell-1) m n (p + \psi^2 + p^2)$$

$$+ \ell(\ell-1) m n (n-1)(\chi^2 + 2p^2) + \ell(\ell-1) m (m-1) n (\psi^2 + p^2)$$

$$+ \ell(\ell-1) m (m-1) n (n-1) p^2 .$$

Hence denoting by σ^2, the variance of W, we have

$$\sigma^2 = E(W^2) - (E(W))^2$$

$$= \ell m n [p + (n-1)(\chi^2 + p^2) + (m-1)(n-1)(\chi^2 + 2p^2) + (m-1)(p + \psi^2 + p^2)$$

$$+ (\ell-1)(p + \psi^2 + p^2) + (\ell-1)(n-1)(\chi^2 + 2p^2)$$

$$+ (\ell-1)(m-1)(\psi^2 + p^2) + (\ell-1)(m-1)(n-1)p^2 - \ell m n p^2]$$

$$= \ell m n [(\ell m - 1)\psi^2 + (n-1)(\ell + m - 1)\chi^2 + (\ell + m - 1)p$$

$$+ \{(n-1)(\ell + m - 2) - 1\}p^2] .$$

Lemma 3. 2

(3.8)
$$\psi^2 \leq p(1-p) , \text{ and}$$
$$\chi^2 \leq p(1-p) .$$

Proof

From (3.6) we have

$$0 \leq \psi^2 = \int_{-\infty}^{\infty} (H(z))^2 dG(z) - p^2$$

$$\leq \int_{-\infty}^{\infty} H(z)\,dG(z) - p^2$$

$$= p - p^2$$

$$= p(1-p) \ .$$

Similarly, from (3.7) we have

$$0 \leq \chi^2 = \int_{-\infty}^{\infty} (1 - G(t))^2 \, dH(t) - p^2$$

$$\leq \int_{-\infty}^{\infty} (1 - G(t) \, dH(t) - p^2$$

$$= 1 - \int_{-\infty}^{\infty} G(t)\,dH(t) - p^2$$

$$= p - p^2$$

$$= p(1-p) \ .$$

Theorem 3.2

$\hat{p} = \dfrac{W}{\ell mn}$ is consistent estimate of p .

Proof

$$Var(\hat{p}) = \frac{1}{\ell^2 m^2 n^2} Var(W)$$

$$= \frac{1}{\ell mn} [(\ell m-1) \ \psi^2 + (n-1)(\ell + m - 1)\chi^2 + (\ell + m - 1)p$$
$$+ \{(n-1)(\ell + m - 2)-1\}p^2]$$

$$\leq \frac{1}{\ell mn} [(\ell m-1)p(1-p) + (n-1)(\ell + m - 1)p(1-p)$$
$$+ (\ell + m - 1)p + \{(n-1)(\ell+m-2)-1\}p^2]$$

$$= \frac{1}{\ell mn} [(\ell m + \ell n + mn - n - 1)p - (\ell m + n - 1)p^2]$$

$$= p(\frac{1}{n} + \frac{1}{m} + \frac{1}{\ell} - \frac{1}{\ell m} - \frac{1}{\ell mn}) - p^2 (\frac{1}{n} + \frac{1}{\ell m} - \frac{1}{\ell mn})$$

$$\to 0 \quad \text{as} \quad \ell, m, n \to \infty \ .$$

3.4 On a Use of a Bivariate Extension of the U Statistic

D. R. Whitney [17] considered a Bivariate extension of the U statistic. Let

U = number of pairs (x_i, z_k) such that $x_i < z_k$,

V = number of pairs (y_j, z_k) such that $y_j < z_k$, where

$$i = 1, \ldots, \ell;$$

$$j = 1, \ldots, m;$$

$$k = 1, \ldots, n.$$

He proposed a test based on the two statistics U and V to test the hypothesis H that $G = F_1 = F_2$ against the alternative that $G > F_1$, $G > F_2$ or say $G > F_1 > F_2$. As a critical region for the hypothesis H with the alternative $G > F_1$, $G > F_2$, he proposed to use $U \leq K_1$, $V \leq K_2$ or with the alternative $G > F_1 > F_2$, $U \geq K_3$, $V \leq K_4$ where the constants K_i are chosen to give the correct significance level. He obtained recurrence relations to determine the probability of a given (U,V) in a sample of ℓx's, my's and nz's. He also evaluated moments of the joint distribution of U and V under the null hypothesis H that $G = F_1 = F_2$ and showed the limit distribution to be Bivariate normal.

Unfortunately, we could not find a suitable statistic based on U and V to estimate the parameter p unbiasedly. However, the statistics U, V and W can be used to estimate the product of marginal probabilities, viz., $p_1 \cdot p_2$ where,

$$p_1 = P\{X < Z\}, \text{ and}$$

$$p_2 = p\{Y < Z\}.$$

The statistics

$$\hat{p}_1 = \frac{U}{\ell n}, \text{ and}$$

$$\hat{p}_2 = \frac{V}{mn}$$

satisfy $E(\hat{p}_1) = p_1$, $E(\hat{p}_2) = p_2$.   Moreover $\hat{p}_1$, $\hat{p}_2$ have other nice properties by themselves, but

$$E(\hat{p}_1\hat{p}_2) \neq p_1p_2$$

since $\hat{p}_1$ and $\hat{p}_2$ are not independent. However, the following theorem gives an unbiased estimate of the product p_1p_2 .

Theorem 3.3

$$\hat{p}_3 = \frac{UV - W}{\ell mn(n-1)}$$

is an unbiased estimate of p_1p_2 .

Proof

First, note that

$$U = \sum_{i=1}^{\ell} \sum_{k=1}^{n} a_{ik} \text{ , and}$$

$$V = \sum_{j=1}^{m} \sum_{k=1}^{n} b_{jk} \text{ ,}$$

so that

$$E(UV) = E\left(\sum_{i=1}^{\ell} \sum_{k=1}^{n} a_{ik} \right)\left(\sum_{j=1}^{m} \sum_{k'=1}^{n} b_{jk'} \right)$$

$$= E\left(\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{k'=1}^{n} (a_{ik} b_{jk'}) \right) ,$$

$$= E\left(\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} a_{ik} b_{jk} + \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{\substack{k'=1 \\ k \neq k'}}^{n} a_{ik} b_{jk'} \right)$$

$$= \ell mnp + \ell mn(n-1)p_1p_2 ,$$

since $E(a_{ik} b_{jk}) = p$ and for $k \neq k'$, a_{ik}, $b_{jk'}$ are independent, so that

$$E(a_{ik} b_{jk'}) = E(a_{ik})E(b_{jk'}) = p_1p_2 .$$

Thus,

$$\hat{p}_3 = \frac{UV - \ell mn\hat{p}}{\ell mn(n-1)}$$

$$= \frac{UV - W}{\ell mn(n-1)}$$

has expected value $p_1 p_2$.

Corollary

If $G = F_1 = F_2$, we have $p = 1/3$, $p_1 = 1/2$ and $p_2 = 1/2$

so that

$$E(UV) = \frac{1}{3} \ell mn + \frac{1}{4} \ell mn(n-1)$$

$$= \frac{1}{12} \ell mn + \frac{1}{4} \ell mn^2 ,$$

which is the same result as derived by D. R. Whitney [17] in Section 3

for the case $G = F_1 = F_2$.

Further properties of $\hat{p}_3$ are not explored, because it does not seem

to be of much importance except perhaps this statistic may be used to

test the hypothesis that all observations come from the same distribution

function against the alternative that the departure from the null hypothe-

sis is in a certain direction.

3.5 A Generalization

Let $X_1,\ldots,X_N,Y$ be (N+1) independent random variables with continuous

cumulative distribution functions $F_1,\ldots,F_N,G$ respectively, and

$$p = \text{Prob}\{X_1 < Y,\ldots,X_N < Y\} .$$

Furthermore, suppose that the samples $(x_{11},\ldots,x_{1m_1}),\ldots,(x_{N_1},\ldots,x_{Nm_N})$,

$(y_1,\ldots,y_n)$ of $X_1,\ldots,X_N,Y$, respectively, are available. Then arranging

the samples in ascending order and defining

$$a_{ij_{(i)}k} = \begin{cases} 1 & \text{if } x_{ij_{(i)}} < z_k & i=1,\ldots,N \\ \\ 0 & \text{otherwise} & k=1,\ldots,n \end{cases}$$

$$j_{(i)}=1,\ldots,m_i$$

$$W = \sum_{i=1}^{N} \sum_{j_{(i)}=1}^{m_i} \sum_{k=1}^{n} (\prod_{\ell=1}^{N} a_{ij_{(\ell)}k}) ,$$

and

$$\hat{p} = \frac{W}{m_1 m_2 \ldots m_N \cdot n} .$$

Then it can be shown that $\hat{p}$ is unbiased consistent estimate of p.

CHAPTER IV

A DUAL PROBLEM

4.1 Introduction

Let $X, Y_1, \ldots, Y_N$ be $(N+1)$ random variables with continuous cumulative distribution functions $F, G_1, \ldots, G_N$ respectively. Then, in certain physical situations, as was pointed out in Chapter I, it might be of interest to estimate the Probability that the random variable X is simultaneously less than the N random variables $Y_1, \ldots, Y_N$ over a time interval $[0, T]$. Thus, let

(4.1) $q = \text{Prob}\{X < Y_1, \ldots, X < Y_N\}$.

This problem is called a dual problem. We shall use the following notations and lemmas throughout this chapter. Let

$$T^* = \min(Y_1, \ldots, Y_N),$$

and suppose it has a c.d.f. H^*. Then we have the following lemmas.

Lemma 4.1

$$P(X < Y_1, \ldots, X < Y_N) = P(X < T^*).$$

Lemma 4.2

$$H^*(t) = 1 - \prod_{i=1}^{N} (1 - G_i(t)).$$

Proof

$$H^*(t) = P(T^* \leq t)$$

$$= P(\min(Y_1, \ldots, Y_N) \leq t)$$

-31-

$$= 1 - P\{\min(Y_1,\ldots,Y_N) > t\}$$

$$= 1 - P\{Y_1 > t, Y_2 > t,\ldots,Y_N > t\}$$

(Since $\min(Y_1,\ldots,Y_N) > t \iff Y_1 > t, Y_2 > t,\ldots,Y_N > t$)

$$= 1 - P(Y_1 > t)P(Y_2 > t) - P(Y_N > t)$$

$$= 1 - \prod_{i=1}^{N} (1 - G_i(t)).$$

In the particular case when $N = 2$, $H^*(t)$ reduces to

$$H^*(t) = G_1(t) + G_2(t) - G_1(t)G_2(t) .$$

Since $H^*(t)$ is a function of continuous c.d.f.'s, $H^*(t)$ itself is continuous, and so the probability of getting ties from the c.d.f. H^* is zero. Moreover the random variable T^* is independent of X.

Next, suppose that $G_1 = G_2 = \ldots = G_N \equiv G$, i.e., the random variables $Y_1,\ldots,Y_N$ have the same continuous c.d.f. Also, G has the density g and X has the continuous c.d.f. F with density f and all these random variables are mutually independent. Then the density function of the random variable T^* is given as

$$h^*(t) = N(1 - G(t))^{N-1} g(t) ,$$

and thus

Lemma 4.3

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x)dx .$$

Proof

$$q = P(X < T^*)$$

$$= E_X P(T^* > X | X)$$

$$= \int_{-\infty}^{\infty} (1 - H^*(x)) dF(x)$$

$$= \int_{-\infty}^{\infty} (1 - G(x))^N f(x) dx \ .$$

Lemma 4.4

If $F = G_1 = \ldots G_N = G$, then $q = \frac{1}{N+1}$.

4.2 Estimation of q When $Y_1,\ldots,Y_N$, X are all Independent Normal.

Let $Y_1,\ldots,Y_N$ all be independent identically distributed normal random variables with mean μ, variance σ^2, and let X be normal with mean ν, variance τ^2 and independent of all the Y's. Then

$$q = P(X < Y_1,\ldots,X < Y_N)$$

$$= P(X-Y_1 < 0,\ldots,X-Y_N < 0)$$

$$= P\left(\frac{X-Y_1 - (\nu-\mu)}{(\sigma^2 + \tau^2)^{1/2}} < \frac{\mu - \nu}{(\sigma^2 + \tau^2)^{1/2}},\ldots,\frac{(X-Y_N)-(\nu-\mu)}{(\sigma^2 + \tau^2)^{1/2}} < \frac{\mu - \nu}{(\sigma^2+\tau^2)^{1/2}}\right)$$

$$= P\left(\max\left\{\frac{X-Y_1 - (\nu-\mu)}{(\sigma^2 + \tau^2)^{1/2}},\ldots,\frac{(X-Y_N) - (\nu-\mu)}{(\sigma^2 + \tau^2)^{1/2}}\right\} < \frac{\mu - \nu}{(\sigma^2 + \tau^2)^{1/2}}\right)$$

$$= F_N(\tilde{H},\rho) \ .$$

where

$$\rho = \frac{\tau^2}{(\sigma^2 + \tau^2)} \ , \text{ and}$$

$$\tilde{H} = \frac{\mu - \nu}{(\sigma^2 + \tau^2)^{1/2}} \ .$$

Since, for all $i = 1,\ldots,N$, the random variables

$$\frac{X - Y_i - (\nu - \mu)}{(\sigma^2 + \tau^2)^{1/2}}$$

are equally correlated standarized normal random variables with correlation coefficient ρ. $F_N(\tilde{H}, \rho)$ can be found using Gupta's table [6]. In case of unknown parameters we can use their maximum likelihood estimates.

4.3 Minimum Variance Unbiased Estimation of q in Some Cases

Let X, Y, Z be independent normal with means μ_x, μ_y, μ_z respectively and some variance σ^2. Then

$$q = P(X < Y, X < Z)$$

$$= P\left(\frac{(X - Y)-(\mu_x - \mu_y)}{(2)^{1/2}\sigma} < \frac{\mu_y - \mu_x}{(2)^{1/2}\sigma}\right.,$$

$$\left.\frac{(X - Z)-(\mu_x - \mu_z)}{(2)^{1/2}\sigma} < \frac{\mu_z - \mu_x}{(2)^{1/2}\sigma}\right).$$

Now, $\left(\dfrac{(X - Y)-(\mu_x - \mu_y)}{(2)^{1/2}\sigma}, \dfrac{(X - Z)-(\mu_x - \mu_z)}{(2)^{1/2}\sigma}\right)$ is bivariate normal with

0 means, unit variances and correlation coefficient equal to 1/2. Thus

$$q = B\left(\frac{\mu_y - \mu_x}{(2)^{1/2}\sigma}, \frac{\mu_z - \mu_x}{(2)^{1/2}\sigma}, \frac{1}{2}\right),$$

where $B(h, k, \rho)$ is given by (2.6).

In case of unknown parameters, we can use their maximum likelihood estimates. Also, for minimum variance unbiased estimates, we shall state the following theorems without proof, since the proof is exactly similar to the proof of theorems 2.1 and 2.2.

Theorem 4.1

If the means μ_y, μ_z and variance σ^2 are known, but μ_x is un-

known and the data $(x_1,\ldots,x_n)$ on X are available, then the m.v.u. estimate of q is given as

$$\hat{q} = B\left(\frac{\mu_y - \bar{x}}{\sigma(2-\frac{1}{n})^{1/2}} \quad , \quad \frac{\mu_z - \bar{x}}{\sigma(2-\frac{1}{n})^{1/2}} \quad , \quad \frac{n-1}{2n-1} \right),$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Moreover, this estimate is asymptotically the same as the maximum likelihood estimate for this case.

Theorem 4.2

If the mean μ_x and variance σ^2 are known but the means μ_y and μ_z are unknown, and the data $(y_1,\ldots,y_\ell)$, $(z_1,\ldots,z_m)$ on Y and Z, respectively, are available, then the m.v.u. estimate of q is given by

$$\hat{q} = B\left(\frac{\bar{y} - \mu_x}{\sigma(2-\frac{1}{\ell})^{1/2}} \quad , \quad \frac{\bar{z} - \mu_x}{\sigma(2-\frac{1}{m})^{1/2}} \quad , \quad \frac{1}{[(2-\frac{1}{\ell})(2-\frac{1}{m})]^{1/2}} \right)$$

where

$$\bar{y} = \frac{1}{\ell} \sum_{j=1}^{\ell} y_j , \text{ and}$$

$$\bar{z} = \frac{1}{m} \sum_{k=1}^{m} z_k .$$

Moreover, as ℓ and m become large, this estimate approaches in limit to the maximum likelihood estimate.

4.4 Estimation of q When X Has a Normal Distribution with Mean μ, Variance σ^2 and $Y_1,\ldots,Y_N$ are All Independent Exponential.

Let $Y_1,\ldots,Y_N$ all be independent and identically distributed with the

density function

$$(4.2) \qquad g(y) = \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and the distribution function

$$(4.3) \qquad G(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1-e^{-x} & \text{if } x > 0 \ . \end{cases}$$

Next, suppose that X is a normal random variable with mean μ and variance σ^2. This might be a practical situation, for example, the electric current supplied to a component is normally distributed with certain mean and variance whereas the capacity of the different transistors inside the component are all exponentially distributed. The component does not fail as long as the current supplied to the component is simultaneously less than the capacity of different transistors, otherwise it fails.

In this situation, we have the following result.

Lemma 4.5

If X is normally distributed with mean μ and variance σ^2 and $Y_1, \ldots, Y_N$ are independent identically distributed with density function (4.2), then

$$(4.4) \qquad q = \Phi(-\frac{\mu}{\sigma}) + \exp(-N\mu + \frac{N^2\sigma^2}{2})(1 - \Phi(N\sigma - \frac{\mu}{\sigma})) \ ,$$

where $\Phi(x)$ is defined by (2.4).

Proof

By lemma 4.3, we have

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x) \, dx$$

$$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} e^{-Nx} f(x)\,dx$$

$$= \int_{-\infty}^{0} \frac{1}{(2\pi)^{1/2}\sigma} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}\,dx + \int_{0}^{\infty} e^{-Nx} \frac{1}{\sigma(2\pi)^{1/2}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}\,dx$$

$$= \int_{-\infty}^{-\mu/\sigma} \phi(y)\,dy + e^{-N\mu + \frac{N^2\sigma^2}{2}} \int_{-\left(\frac{\mu-N\sigma^2}{\sigma}\right)}^{\infty} \phi(y)\,dy$$

$$= \Phi(-\mu/\sigma) + e^{-N\mu + \frac{N^2\sigma^2}{2}}\left(1 - \Phi\left(N\sigma - \frac{\mu}{\sigma}\right)\right).$$

Corollary: If $\mu = 0$, $\sigma = 1$, then

$$q = \frac{1}{2} + e^{N^2/2}(1 - \Phi(N)).$$

Now, suppose that a random sample distributed as X, say $x_1, \ldots, x_m$ is observed. The problem here is to find the minimum variance unbiased estimate of the q by use of this sample. Letting $\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$, this estimate is provided by:

Theorem 4.3

If σ is known and $m > 2$, the minimum variance unbiased estimate of q is

$$(4.5) \quad \hat{q} = \Phi\left(\frac{-(m)^{1/2}\bar{x}}{\sigma(m-1)^{1/2}}\right) + e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}}$$

$$\cdot \left[1 - \Phi\left\{\frac{(m-1)^{3/2}N\sigma^2 - m(m-1)^{1/2}\bar{x}}{\sigma(m)^{1/2}[1 + m(m-2)]^{1/2}}\right\}\right] \quad .$$

And as m becomes large, this estimate approaches in the limit to the maximum likelihood estimate of q.

Proof

Let $q = q_1 + q_2$,

where $q_1 = \Phi(-\mu/\sigma)$, and

$$q_2 = e^{-N\mu + \frac{N^2\sigma^2}{2}}(1 - \Phi(N\sigma - \frac{\mu}{\sigma})).$$

To find the m.v.u. estimate of q, it would be enough to find an unbiased estimate of q which is a function of a complete-sufficient statistic, in view of the Rao-Blackwell and Lehman-Scheffe´ theorems. A complete-sufficient statistic in this case is $\bar{x}$ and if we can find an unbiased estimate $\hat{q}_1$ of q_1 and an unbiased estimate $\hat{q}_2$ of q_2 which are functions of a complete-sufficient statistic, then m.v.u. estimate of q will be

$$\hat{q} = \hat{q}_1 + \hat{q}_2 \quad .$$

Note that

$$q_1 = \int_{-\infty}^{0} \frac{1}{(2\pi)^{1/2}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \quad ,$$

so that, following Lieberman and Reonikoff [8], the m.v.u. estimate of q_1 is given by

$$\hat{q}_1 = \Phi(-\frac{(m)^{1/2}\bar{x}}{(m-1)^{1/2}\sigma}) \quad ,$$

which is a function of a complete sufficient statistic.

Thus, we only need to find an unbiased estimate of q_2 which is a function of a complete sufficient statistic.

Let $t_1 = e^{-Nx_1}$.

Then,

$$E(t_1) = \int e^{-Nx_1} \frac{1}{(2\pi)^{1/2}\sigma} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} dx_1$$

$$= \int \frac{1}{(2\pi)^{1/2}\sigma} e^{-\frac{1}{2\sigma^2}(x_1-(\mu-N\sigma^2))^2} dx_1 \; e^{-N\mu + \frac{N^2\sigma^2}{2}}$$

$$= e^{-N\mu + \frac{N^2\sigma^2}{2}}$$

Further, let

$$t_2 = \begin{cases} 0 & \text{if } -\infty < x_2 < N\sigma^2 \\ 1 & \text{otherwise ,} \end{cases}$$

then

$$E(t_2) = 1 - \text{Prob}(-\infty < x_2 < N\sigma^2)$$

$$= 1 - \int_{-\infty}^{N\sigma - \frac{\mu}{\sigma}} \phi(y)dy$$

$$= 1 - \Phi(N\sigma - \frac{\mu}{\sigma}) \; .$$

Since t_1 is a function of x_1 alone and t_2 is a function of x_2 alone, x_1, x_2 being independent, implies that t_1 and t_2 are independent. Thus, let

$$t = t_1 t_2 \; , \text{ and}$$

we obtain

$$E(t) = E(t_1)E(t_2)$$

$$= e^{-N\mu + \frac{N^2\sigma^2}{2}} (1 - \Phi(N\sigma - \frac{\mu}{\sigma}))$$

$$= q_2 \; .$$

Thus, in view of Rao-Blackwell, Lehmann-Scheffe' theorems, since $\bar{x}$ is a complete-sufficient statistic, $E(t|\bar{x})$ is m.v.u. estimate of q_2. Note that

$$t = \begin{cases} e^{-Nx_1} & \text{if } N\sigma^2 < x_2 < \infty, \text{ and } -\infty < x_1 < \infty \\ 0 & \text{otherwise} \end{cases}$$

$$= e^{-Nx_1} I_{(N\sigma^2,\infty)}(x_2) \qquad\qquad -\infty < x_1 < \infty \; ,$$

where

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a < x < b \\ 0 & \text{otherwise .} \end{cases}$$

Thus,

$$E(t|\bar{x}) = E(e^{-Nx_1} I_{(N\sigma^2,\infty)}(x_2)|\bar{x})$$

Now,

$$\begin{pmatrix} \bar{x} \\ x_1 \\ x_2 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \dfrac{\sigma^2}{m} & \dfrac{\sigma^2}{m} & \dfrac{\sigma^2}{m} \\ \dfrac{\sigma^2}{m} & \sigma^2 & 0 \\ \dfrac{\sigma^2}{m} & 0 & \sigma^2 \end{pmatrix} \right) .$$

Thus, using 8a.2(v) Page 441, Rao [14], the conditional distribution of (x_1, x_2) given $\bar{x}$ is

$$N_2 \left(\begin{pmatrix} \bar{x} \\ \bar{x} \end{pmatrix}, \begin{pmatrix} (1 - \dfrac{1}{m})\sigma^2 & -\dfrac{\sigma^2}{m} \\ -\dfrac{\sigma^2}{m} & (1 - \dfrac{1}{m})\sigma^2 \end{pmatrix} \right),$$

and the conditional correlation coefficient is $-\dfrac{1}{(m-1)}$. Therefore, the conditional density of x_1, x_2 given $\bar{x}$, after simplification, reduces to

$$f(x_1,x_2|\bar{x}) = (\tfrac{m}{m-2})^{1/2} \frac{1}{2\pi\sigma^2} e^{-\frac{m}{2(m-2)\sigma^2}[(\bar{x}-x_1)^2+(\bar{x}-x_2)^2 - \frac{(x_1-x_2)^2}{m}]} ,$$

Hence,

$$E(t|\bar{x}) = E(e^{-Nx_1} I_{(N\sigma^2,\infty)}(x_2)|\bar{x})$$

$$= \int_{-\infty}^{\infty} \int_{N\sigma^2}^{\infty} e^{-Nx_1} (\tfrac{m}{m-2})^{1/2} \frac{1}{2\pi\sigma^2} e^{-\frac{m}{2(m-2)\sigma^2}((\bar{x}-x_1)^2+(\bar{x}-x_2)^2 - \frac{1}{m}(x_1-x_2)^2)}$$

$$\cdot\ dx_2\, dx_1$$

$$= \int_{-\infty}^{\infty} e^{-Nx_1 - \frac{m}{2(m-2)\sigma^2}(\bar{x}-x_1)^2} (\tfrac{m}{m-2})^{1/2} \frac{1}{2\pi\sigma^2} \int_{N\sigma^2}^{\infty} e^{-\frac{m}{2(m-2)\sigma^2}((\bar{x}-x_2)^2 -}$$

$$-\frac{1}{m}(x_1-x_2)^2)\ dx_2\, dx_1$$

The exponent term, except for the constant, of the second integral, is

$$(\bar{x} - x_2)^2 - \frac{1}{m}(x_1 - x_2)^2 = \bar{x}^2 + x_2^2 - 2x_2\bar{x} - \frac{1}{m}(x_1^2 + x_2^2 - 2x_1x_2)$$

$$= \bar{x}^2 - \frac{1}{m}x_1^2 + \frac{m-1}{m}[x_2 - \frac{m}{m-1}(\bar{x} - \frac{1}{m}x_1)]^2$$

$$- \frac{m}{m-1}(\bar{x} - \frac{1}{m}x_1)^2 .$$

So that,

$$\frac{(m-1)^{1/2}}{(2\pi)^{1/2}\sigma(m-2)^{1/2}} \int_{N\sigma^2}^{\infty} e^{-\frac{m}{2(m-2)\sigma^2}\frac{m-1}{m}(x_2 - \frac{m}{m-1}(\bar{x} - \frac{1}{m}x_1))^2} dx_2$$

$$= \int_{k\{N\sigma^2 - \frac{m}{m-1}(\bar{x} - \frac{1}{m}x_1)\}}^{\infty} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}t^2} dt$$

(where $1/k = (\frac{m-2}{m-1})^{1/2}\sigma$)

$$= 1 - \Phi\{kN\sigma^2 - \frac{mk}{m-1}(\bar{x} - \frac{1}{m}x_1)\}.$$

Thus

$$E(t|\bar{x}) = \int_{-\infty}^{\infty} e^{-Nx_1 - \frac{m}{2(m-2)\sigma^2}(\bar{x}-x_1)^2 - \frac{m}{2(m-2)\sigma^2}(\bar{x}^2 - \frac{1}{m}x_1^2 - \frac{m}{m-1}(\bar{x} - \frac{1}{m}x_1)^2)}$$

$$\cdot \frac{(m)^{1/2}}{(m-1)^{1/2}} \frac{1}{(2\pi)^{1/2}\sigma} [1 - \Phi(Nk\sigma^2 - \frac{mk}{m-1}(\bar{x} - \frac{1}{m}x_1))]dx_1 .$$

Simplifying the exponent term again

$$-Nx_1 - \frac{m}{2(m-2)\sigma^2}[\bar{x}^2 + x_1^2 - 2x_1\bar{x} + \bar{x}^2 - \frac{1}{m}x_1^2 - \frac{m}{m-1}(\bar{x}^2 + \frac{1}{m^2}x_1^2 - \frac{2}{m}x_1\bar{x})]$$

$$= -Nx_1 - \frac{m}{2(m-2)\sigma^2} \frac{1}{m-1}[2(m-1)\bar{x}^2 + (m-1)(1 - \frac{1}{m})x_1^2 - 2(m-1)x_1\bar{x} - m\bar{x}^2$$

$$- \frac{1}{m}x_1^2 + 2x_1\bar{x}]$$

$$= -Nx_1 - \frac{m}{2(m-2)\sigma^2} \frac{1}{m-1}[(m-2)\bar{x}^2 + \frac{(m-1)^2 - 1}{m}x_1^2 - 2(m-2)x_1\bar{x}]$$

$$= -Nx_1 - \frac{m}{2(m-1)\sigma^2}[\bar{x}^2 + x_1^2 - 2x_1\bar{x}]$$

$$= - \frac{m}{2(m-1)\sigma^2}(x_1^2 - 2(\bar{x} - \frac{N(m-1)\sigma^2}{m})x_1 + \bar{x}^2)$$

$$= - \frac{m}{2(m-1)\sigma^2}[(x_1 - (\bar{x} - \frac{N(m-1)\sigma^2}{m}))^2 - (\bar{x} - \frac{N(m-1)\sigma^2}{m})^2 + \bar{x}^2]$$

$$= - \frac{m}{2(m-1)\sigma^2}[x_1 - (\bar{x} - \frac{N(m-1)\sigma^2}{m})]^2 - N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m} \ .$$

Therefore,

$$E(t|\bar{x}) = e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}}$$

$$\cdot \left[\int_{-\infty}^{\infty} \frac{(m)^{1/2}}{(m-1)^{1/2}} \frac{1}{(2\pi)^{1/2}\sigma} e^{- \frac{m}{2(m-1)\sigma^2}[x_1-(\bar{x} - \frac{N(m-1)\sigma^2}{m})]^2} \right.$$

$$\left. \{1 - \Phi(Nk\sigma^2 - \frac{mk}{m-1}(\bar{x} - \frac{1}{m} x_1))\}dx_1 \right]$$

$$= e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}}$$

$$\cdot \int_{-\infty}^{\infty} \frac{(m)^{1/2}}{(m-1)^{1/2}} \frac{1}{(2\pi)^{1/2}\sigma} e^{- \frac{m}{2(m-1)\sigma^2}[x_1-(\bar{x} - \frac{N(m-1)\sigma^2}{m})]^2} dx_1$$

$$- \int_{-\infty}^{\infty} \frac{(m)^{1/2}}{(m-1)^{1/2}} \frac{1}{(2\pi)^{1/2}\sigma} \Phi(Nk\sigma^2 - \frac{mk}{m-1} (\bar{x} - \frac{1}{m} x_1))$$

$$\cdot e^{- \frac{m}{2(m-1)\sigma^2}\{x_1- (\bar{x} - \frac{N(m-1)\sigma^2}{m})\}^2}dx_1$$

$$= e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}}[1 - \int_{-\infty}^{\infty} \frac{(m)^{1/2}}{(m-1)^{1/2}} \frac{1}{(2\pi)^{1/2}\sigma} \Phi\{Nk\sigma^2 - \frac{mk}{m-1} \cdot$$

$$\cdot (\bar{x} - \frac{1}{m} x_1)\}e^{- \frac{m}{2(m-1)\sigma^2}\{x_1-(\bar{x}- \frac{N(m-1)\sigma^2}{m})\}^2} dx_1] \ .$$

To evaluate the integral, use the transformation

$$t = \frac{x_1 - (\bar{x} - \frac{N(m-1)\sigma^2}{m})}{(\frac{m-1}{m})^{1/2}\sigma} \quad , \text{ so that}$$

$$x_1 = (\frac{m-1}{m})^{1/2}\sigma t + \bar{x} - \frac{N(m-1)\sigma^2}{m} ,$$

and the jacobian transformation is

$$dx_1 = (\frac{m-1}{m})^{1/2}\sigma dt.$$

Thus the integral becomes

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \Phi[Nk\sigma^2 - \frac{mk}{m-1} \{\bar{x} - \frac{1}{m}(\frac{m-1}{m})^{1/2}\sigma t - \frac{\bar{x}}{m} + \frac{N(m-1)\sigma^2}{m^2}\}]e^{-\frac{1}{2}t^2} dt$$

$$= \int_{-\infty}^{\infty} \Phi[Nk\sigma^2 - \frac{k}{m-1} \{(m-1)\bar{x} - (\frac{m-1}{m})^{1/2}\sigma t + \frac{N(m-1)\sigma^2}{m}\}]\phi(t) dt$$

$$= \int_{-\infty}^{\infty} \Phi[Nk\sigma^2 - k\bar{x} + \frac{k}{[m(m-1)]^{1/2}} \sigma t - \frac{Nk}{m} \sigma^2]\phi(t) dt$$

$$= \int_{-\infty}^{\infty} \phi(t)\Phi[(1 - \frac{1}{m})Nk\sigma^2 - k\bar{x} + \frac{k\sigma}{[m(m-1)]^{1/2}} t] dt$$

$$= \int_{-\infty}^{\infty} \phi(t)\Phi(a + bt) dt$$

$$= \Phi\left(\frac{a}{(1 + b^2)^{1/2}}\right) ,$$

following 7.6.2, Page 35 of Owen [11] where

$$k = \frac{(m-1)^{1/2}}{(m-2)^{1/2}\sigma} ,$$

$$a = (1 - \frac{1}{m})Nk\sigma^2 - k\bar{x}$$

$$= \frac{m-1}{m} \frac{(m-1)^{1/2}}{(m-2)^{1/2}} N\sigma - (\frac{m-1}{m-2})^{1/2} \frac{\bar{x}}{\sigma} ,$$

$$b = \frac{k}{[m(m-1)]^{1/2}}$$

$$= \frac{(m-1)^{1/2}}{(m-2)^{1/2}[m(m-1)]^{1/2}} = \frac{1}{[m(m-2)]^{1/2}} .$$

Therefore,

$$\hat{q}_2 = E(t|\bar{x}) = e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}} [1 - \Phi(\frac{a}{(1 + b^2)^{1/2}})]$$

$$= e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}} \left[1 - \Phi \left\{ \frac{\frac{(m-1)^{3/2}}{m(m-2)^{1/2}} N\sigma - (\frac{m-1}{m-2})^{1/2} \frac{\bar{x}}{\sigma}}{[1 + \frac{1}{m(m-2)}]^{1/2}} \right\} \right]$$

$$= e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}} [1 - \Phi \left\{ \frac{(m-1)^{3/2}N\sigma^2 - m(m-1)^{1/2}\bar{x}}{\sigma(m)^{1/2}[m(m-2) + 1]^{1/2}} \right\}] ,$$

and the minimum variance unbiased estimate of q is, therefore,

$$\hat{q} = \hat{q}_1 + \hat{q}_2$$

$$= \Phi(-\frac{(m)^{1/2}\bar{x}}{\sigma(m-1)^{1/2}}) + e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}} [1 - \Phi \left\{ \frac{(m-1)^{3/2}N\sigma^2 - m(m-1)^{1/2}\bar{x}}{\sigma(m)^{1/2}[1 + m(m-2)]^{1/2}} \right\}] .$$

The maximum likelihood estimate of q when σ is known is given as

$$\hat{q} = \Phi(-\frac{\bar{x}}{\sigma}) + e^{-N\bar{x} + \frac{N^2\sigma^2}{2}} [1 - \Phi(N\sigma - \frac{\bar{x}}{\sigma})] ,$$

and thus as m becomes large, the minimum variance unbiased estimate is asymptotically the same as the maximum likelihood estimate.

Corollary 4.1

Let $Y_1, \ldots, Y_N$ all be independent exponential random variables with known parameter λ, i.e., the density of Y's is given as

$$f(y) = \begin{cases} \alpha \, e^{-\alpha y} & y > 0, \ \alpha > 0 \\ 0 & \text{otherwise} . \end{cases}$$

Also let X be an independent normal random variable with unknown mean μ and known variance σ^2. Then, the minimum variance unbiased estimate of q based on the random sample $x_1, \ldots, x_m$ from X, is

$$\hat{q} = \Phi\left[\frac{-(m)^{1/2}\bar{x}}{\sigma(m-1)^{1/2}}\right] + e^{-\alpha N\bar{x} + \frac{\alpha^2 N^2 (m-1)\sigma^2}{2m}}\left[1-\Phi\left\{\frac{(m-1)^{3/2}\alpha N\sigma^2 - m(m-1)^{1/2}\bar{x}}{\sigma(m)^{1/2}[1+m(m-2)]^{1/2}}\right\}\right] .$$

Proof

Note that the c.d.f. of Y is

$$G(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1-e^{-\alpha x} & \text{if } x > 0 . \end{cases}$$

Thus,

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x)\,dx$$

$$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} e^{-\alpha Nx} f(x)\,dx$$

$$= \Phi(-\tfrac{\mu}{\sigma}) + e^{-\alpha N\mu + \frac{\alpha^2 N^2 \sigma^2}{2}} [1 - \Phi(\alpha N\sigma - \tfrac{\mu}{\sigma})], \text{ and therefore}$$

by replaceing N by αN in lemma 4.5 and in theorem 4.3, we get the desired result.

Note that if μ and σ are both unknown in lemma 4.5, then the maximum likelihood estimate of q is

$$\tilde{q} = \Phi(-\tfrac{\bar{x}}{s}) + e^{-N\bar{x} + \frac{N^2 s^2}{2}} (1 - \Phi(Ns - \tfrac{\bar{x}}{s})),$$

where $\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$, and

$$s^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \bar{x})^2 .$$

Also, if the parameter α is unknown in Cor. 4.1 and the data $y_{11}, \ldots, y_{1m_1}, \ldots, y_{N1}, \ldots, y_{Nm_N}$ are available, then the maximum likelihood estimate of q is

$$\hat{q} = \Phi(-\frac{\bar{x}}{s}) + e^{-\bar{y}N\bar{x} + \frac{\bar{y}^2 N^2 s^2}{2}} [1 - \Phi(\bar{y}Ns - \frac{\bar{x}}{s})] ,$$

where
$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} \bar{y}_i ,$$

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij} \qquad i=1,\ldots,N,$$

$\bar{x}$ and s^2 are defined above.

4.5 Estimation of q when X is uniform and $Y_1,\ldots,Y_N$ are all Independent Exponential

Let $Y_1,\ldots,Y_N$ be all independent and identically distributed with the density function and distribution function given by (4.2) and (4.3) respectively, and X be uniform random variable between 0 and θ, i.e., density function of X is given by

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 < x < \theta \\ 0 & \text{otherwise} , \end{cases}$$

then we have the following lemma.

Lemma 4.6

In above case

$$q = \frac{1 - e^{-N\theta}}{N\theta} \; .$$

Proof

By lemma 4.3, we have

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x) \, dx$$

$$= \int_{0}^{\theta} \frac{e^{-Nx}}{\theta} \, dx$$

$$= \frac{1 - e^{-N\theta}}{N\theta} \; .$$

Theorem 4.4

In this case the minimum variance unbiased estimate of q is

$$\hat{q} = \frac{m-1}{mNt} + \frac{1}{m} e^{-Nt}(1 - \frac{m-1}{Nt}) \, ,$$

where $t = \max(x_1, \ldots, x_m)$ and $x_1, \ldots, x_m$ is observed random sample distributed as X.

Proof

Let $u = e^{-Nx_1}$, then we obtain

$$E(u) = q \, .$$

Thus, since t is a complete sufficient statistic, in view of Rao-Blackwell and Lehmann-Scheffe´ theorems, $E(u|t)$ is the m.v.u. estimate of q. The conditional distribution of x_1 given t is of mixed type and its generalized probability density function with respect to mixture of Lebesgue and counting measure is, as given by Patil and Wani [13],

$$h(x_1|t) = \begin{cases} (1 - \frac{1}{m}) \frac{1}{t} & , \quad 0 < x_1 < t \\[2mm] \frac{1}{m} & \quad x_1 = t \\[2mm] 0 & \quad \text{otherwise.} \end{cases}$$

Thus,

$$E(u|t) = \int_0^t e^{-Nx_1}(1 - \frac{1}{m}) \frac{1}{t} \, dx_1 + \frac{1}{m} e^{-Nt}$$

$$= \frac{1}{m} e^{-Nt} + \frac{m-1}{mt}(\frac{1-e^{-Nt}}{N})$$

$$= \frac{m-1}{mNt} + \frac{1}{m} e^{-Nt} (1 - \frac{m-1}{Nt}).$$

4.6 Nonparametric Estimation

Let X, Y, Z be independent random variables with continuous cumulative distribution functions F, G_1, G_2 respectively, and

$$q = \text{Prob}(X < Y \quad \text{and} \quad X < Z)$$
$$= \text{Prob}(X < T^*),$$

where $T^* = \min(Y,Z)$ and let H^* be c.d.f. of T^*.

Further, assume that

$$x_1,\ldots,x_\ell \ , \quad y_1,\ldots,y_m \ , \quad z_1,\ldots,z_n$$

are samples of X, Y and Z respectively. Then, the problem of estimating q on the basis of samples of X, Y and Z is exactly similar to that of estimating $p = \text{Prob}(X < Z \ , \ Y < Z)$ discussed in Chapter III. However, the main results are stated without proof in the following, since the proof follows exactly in the same way as in Chapter III.

Let
$$a_{ij} = \begin{cases} 1 & \text{if } x_i < y_j \\ 0 & \text{otherwise ,} \end{cases}$$

$$b_{ik} = \begin{cases} 1 & \text{if } x_i < z_k \\ 0 & \text{otherwise, where} \end{cases}$$

$$i = 1,\dots,\ell \quad ; \quad j = 1,\dots,m \quad ; \quad k = 1,\dots,n \quad.$$

Further, suppose that

$$U_{ijk} = a_{ij}b_{ik} \text{ , and}$$

$$W_1 = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} U_{ijk}$$

= number of triplets (x_i, y_j, z_k) such that $x_i < y_j$,
$x_i < z_k$,

i.e., W_1 counts the number of times an x precedes a y and a z simultaneously in the arranged sample of ℓx's, my's and nz's in ascending order. Then we have the following results.

Lemma 4.7

$\hat{q} = \dfrac{W_1}{\ell mn}$ is an unbiased estimate of q.

Proof

$U_{ijk} = 1$ if and only if $x_i < y_j$ and $x_i < z_k$ so that $E(U_{ijk}) = P(x_i < y_j , x_i < z_k) = q$, and the proof follows.

Example 4.1

Let $\ell = 4$, m = 3, n = 2, and consider the observations

$$x_2 < y_3 < z_2 < x_1 < x_4 < y_1 < z_1 < x_3 < y_2.$$

There are 3 values of x which are less than (y_1, z_1),

there is 1 value of x which is less than (y_1, z_2),

there are 3 values of x which are less than (y_2, z_1),

there is 1 value of x which is less than (y_2, z_2),

there is 1 value of x which is less than (y_3, z_1), and

there is 1 value of x which is less than (y_3, z_2).

Thus the experimental value of W_1 is

$$w_1 = 3 + 1 + 3 + 1 + 1 + 1 = 10 ,$$

and $\hat{q} = \frac{10}{24}$.

Also,

$a_{11} = 1, a_{21} = 1, a_{31} = 0, a_{41} = 1, a_{12} = 1, a_{22} = 1, a_{32} = 1,$

$a_{42} = 1, a_{13} = 0, a_{23} = 1, a_{33} = 1, a_{43} = 0,$

and

$b_{11} = 1, b_{21} = 1, b_{31} = 0, b_{41} = 1, b_{12} = 0, b_{22} = 1, b_{32} = 0$

$b_{42} = 0.$

Thus,

$$\hat{q} = \frac{1}{\ell mn} \sum_i \sum_j \sum_k a_{ij} b_{ik} = \frac{10}{24} .$$

Next, suppose that

$$\phi^2 = \sigma^2_{F(T*)} = \int_{-\infty}^{\infty} (F(t) - q)^2 dH*(t), \text{ and}$$

$$\nu^2 = \sigma^2_{H*(x)} = \int_{-\infty}^{\infty} (H*(x) - 1 + q)^2 dF(x) .$$

Then, we have the following relations.

$$q = \int_{-\infty}^{\infty} F(t) dH*(t) ,$$

$$1-q = \int_{-\infty}^{\infty} H*(x) dF(x) ,$$

$$\int_{-\infty}^{\infty} F^2(t)\,dH^*(t) = q^2 + \phi^2 \ , \text{ and}$$

$$\int_{-\infty}^{\infty} (1 - H^*(x))^2\,dF(x) = \nu^2 + q^2 \ .$$

Thus, we have the following results.

Theorem 4.5

The variance of W_1 is given by

$$\text{Var}(W_1) = \ell mn[(mn-1)\nu^2 + (\ell-1)(m + n - 1)\phi^2 + (m + n - 1)q$$
$$+ \{(\ell-1)(m + n - 2) - 1\}q^2].$$

Lemma 4.8

$\phi^2 \leq q(1 - q)$, and

$\nu^2 \leq q(1 - q)$.

Theorem 4.6

$\hat{q}$ is consistent estimate of q.

Proof

$$\text{Var}(\hat{q}) = \frac{1}{\ell^2 m^2 n^2}\ \text{Var}(W_1)$$

$$= \frac{1}{\ell mn}[(mn-1)\nu^2 + (\ell-1)(m + n - 1)\phi^2 + (m + n - 1)q$$
$$+ \{(\ell-1)(m + n - 2) - 1\}q^2]$$

$$\leq q\left(\frac{1}{m} + \frac{1}{n} - \frac{1}{mn} + \frac{1}{\ell} - \frac{1}{\ell mn}\right) + q^2\left(\frac{1}{\ell} + \frac{1}{mn} - \frac{1}{\ell mn}\right)$$

$$\text{(using lemma 4.8)}$$

$$\to 0 \text{ as } \ell, m, n \to \infty \ .$$

The above results can also be generalized.

CHAPTER V

CONCLUSION AND FURTHER RESEARCH

The major goal of this dissertation is to discuss estimation of
the reliability of a component subject to several different stresses or
resistances over a time interval $[0,T]$, i.e., the problem is to provide
the estimate of the probability that N random variables are simultaneously
less than a random variable, where all these random variables are mutually
independent. This probability has been called p, and its dual probability,
q, as defined in Chapter IV. Parametric and nonparametric methods of esti-
mation were considered.

In a very few cases, it was possible to derive the minimum variance
unbiased estimates, whereas in other cases the derivation of the mini-
mum variance unbiased estimates remain an open question. For example,
in Theorem 4.3, if σ is also unknown, which is the more realistic case,
we could not derive the minimum variance unbiased estimate. The problem
lies in finding the m.v.u. estimate of $F(c\sigma^2)$, where c is a constant,
and F is cumulative distribution function of a normal variate with mean
μ, and variance σ^2. More generally, if x is a function of μ and σ^2,
then the m.v.u. estimate of $F(x)$ has not yet been found to the best of
our knowledge except when σ^2 is known.

The statistic W based on ranks of the pooled sample was suggested
to estimate p unbiasedly and was shown to have certain optimum proper-

ties. It is felt that the statistic W is of much importance, because only the assumption of the continuity of the cumulative distribution function is needed. However, the results could be extended to the discontinuous case with slight modifications. It is not easy to compute W especially when the number of observations available on each random variable is large, even in the case of N(number of random variables) = 2.

Though the use of the statistic W has been limited only to estimate p in this dissertation, it could be used to test the hypothesis that the random variables $X_1, \ldots, X_N, Y$ have the same continuous cumulative distribution function against the alternative that Y is stochastically larger than the N random variables $X_1, \ldots, X_N$ simultaneously, i.e., Y is stochastically larger than the maximum of $X_1, \ldots, X_N$. This particular use of the statistic W and possibly some other uses are open for further research. We are presently investigating this particular use of the W statistic for the special case N = 2, and as a critical region for the hypothesis, it is proposed to use $W \leq K$, where K is chosen to give the correct significance level. Properties of this test function, the distribution of the statistic W under the null hypothesis, approximation to the normality, a comparison with Whitney's [17] result, etc. are under consideration. Properties of this test function for the general case and other uses of the statistic W are yet to be explored.

REFERENCES

1. Birnbaum, Z. W. (1956), "On a Use of the Mann-Whitney Statistic," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, (University of California Press, Berkeley, California), Vol. 1, pp. 13-17.

2. Birnbaum, Z. W. and McCarty, R. C. (1958), "A Distribution Free Upper Bound for $Pr(Y < X)$ Based on Independent Samples of X and Y," Ann. Math. Stat., Vol. 29, pp. 558-562.

3. Church, J. D. and Bernard Harris (1970), "The Estimation of Reliability from Stress-Strength Relationship," Technometrics, Vol. 12, pp. 49-54.

4. VanDantzig, D. (1951), "On the Consistency and Power of the Wilcoxon's Two-Sample Test," Knonik. Neder. Akad. Van Wetensch. Proc. Ser A., Vol. 54, pp. 1-9.

5. Govindarajulu, Z. (1968), "Distribution-Free Confidence Bounds for $P(X < Y)$," Annals of the Institute of Statistical Mathematics, Vol. 20, pp. 229-238.

6. Gupta, S. S. (1963), "Probability Integrals of Multivariate Normal and Multivariate t," Ann. Math. Stat., Vol. 34, pp. 792-828.

7. Lehmann, E., Notes on the Theory of Estimation, Mimeo notes (University of California Press, Berkeley)(Reprinted 1962).

8. Lieberman, G. J. and Resnikoff, G. J. (1955), "Sampling Plans for Inspection by Variables," Journal of the American Statistical Association, Vol. 50, pp. 457-516.

9. Mann, H. B. and Whitney, D. R. (1947), "On a Test Whether One of Two Random Variables is Stochastically Larger Than the Other," Ann. Math. Stat., Vol. 18, pp. 50-60.

10. Mazumdar, M. (1970), "Some Estimates of Reliability Using Interference Theory," Naval Research Logistics Quarterly, Vol. 17, pp. 159-166.

11. Owen, D. B. (1957), "The Bivariate Normal Probability Distribution," Research Report, SC-3831(TR) Systems Analysis, Sandia Corporation.

12. Owen, D. B., K. J. Craswell and D. L. Hanson (1964), "Nonparametric Upper Confidence Bounds for $P(Y < X)$ and Confidence Limits for $P(Y < X)$ when X and Y are Normal," *Journal of the American Statistical Association*, Vol. 59, pp. 906-924.

13. Patil, G. P. and Wani, J. K. (1966), "Minimum Variance Unbiased Estimation of the Distribution Function Admitting a Sufficient Statistic," *Ann. Inst. Stat. Math.*, Vol. 18, pp. 39-47.

14. Rao, C. R. (1965), *Linear Statistical Inference and its Applications*, John Wiley and Sons, Inc.

15. Sen, P. K. (1960), "On Some Convergence properties of U-Statistic," *Calcutta Stat. Bull.*, Vol. 10, pp. 1-18.

16. Sen, P. K. (1967), "A Note on Asymptotically Distribution-Free Confidence Bounds for $P(X < Y)$ Based on Two Independent Samples," *Sankhyā*, Series A, Vol. 29, pp. 95-102.

17. Whitney, D. R. (1951), "A Bivariate Extension of the U Statistic," *Ann. Math. Stat.*, Vol. 22, pp. 274-282.

18. Zehna, P. W. (1966), "Invariance of Maximum Likelihood Estimation," *Ann. Math. Stat.*, Vol. 37, p. 755.

END